

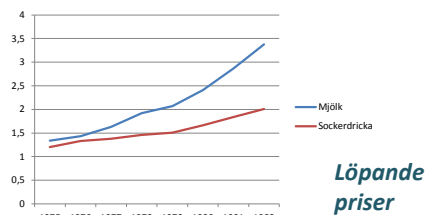
F3-5

Statistikens grunder 2 dagtid

HT 2012

F3 Lite till om tidsserier

- **Deflatering**, att justera för inflationen



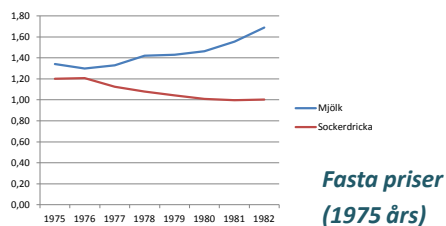
År	Mjolk	Sockerdricka	KPI (1945 = 100)	← SCB
1975	1,34	1,20	347	
1976	1,43	1,33	382	
1977	1,63	1,38	426	
1978	1,92	1,46	469	
1979	2,07	1,51	502	
1980	2,41	1,66	571	
1981	2,87	1,84	640	
1982	3,38	2,01	695	

Deflatering

År	Löpande priser		KPI		Fasta priser	
	Mjolk	Sockerdricka	(1945 = 100)	(1975 = 100)	Mjolk	Sockerdricka
1975	1,34	1,20	347	100,0	1,34	1,20
1976	1,43	1,33	382	110,1	1,30	1,21
1977	1,63	1,38	426	122,8	1,33	1,12
1978	1,92	1,46	469	135,2	1,42	1,08
1979	2,07	1,51	502	144,7	1,43	1,04
1980	2,41	1,66	571	164,6	1,46	1,01
1981	2,87	1,84	640	184,4	1,56	1,00
1982	3,38	2,01	695	200,3	1,69	1,00

$$I_{\text{år}}^{1975} = 100 \cdot \frac{I_{\text{år}}^{1945}}{I_{1975}^{1945}} = 100 \cdot \frac{I_{\text{år}}^{1945}}{347}$$

(Fast pris) $_{\text{år}}^{1975} = 100 \cdot \frac{\text{Pris}_{\text{år}}}{I_{\text{år}}^{1975}}$



Sammansatta index 4

- **Laspeyres index:**

– Kvantiteter vid basåret = 0

$$\text{Index}_t^L = 100 \cdot \frac{\sum P_t Q_0}{\sum P_0 Q_0}$$

- **Paasche index:**

– Kvantiteter innevarande år t

$$\text{Index}_t^P = 100 \cdot \frac{\sum P_t Q_t}{\sum P_0 Q_t}$$

Inlämningsuppgift!

Samplingfördelningar 1

- Nyquist Kap 15
- Tidigare: mycket fokus på en stokastisk variabel och dess fördelning
- Vi har även tittat på simultana fördelningar, två eller tre s.v.
- Nu: ta ett helt urval och titta på urvalets egenskaper.
- Senare: vi ska använda urvalet för att dra allmänna slutsatser (empirism)

Samplingfördelningar 2

- Ett **urval** eller med andra ord **stickprov** är en uppsättning s.v. som betecknas med versaler:

$$X_1, X_2, \dots, X_n$$

- När de har **observerats** betecknas de med gemener:

$$x_1, x_2, \dots, x_n$$

- Ofta skriver man ett enkelt **s** för att beteckna stickprovet

$$s = \{x_1, x_2, \dots, x_n\}$$

Samplingfördelningar 3

- **Egenskaper** hos stickprovet kan bl.a. vara:
 - Stickprovsmedelvärde $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$
(Brukar kallas *x-bar*)
 - Stickprovsvariansen

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$
 - Andelen i stickprovet med ngn egenskap

$$p = \frac{1}{n} \sum_{i=1}^n x_i \quad x_i = \begin{cases} 1 & \text{om i ha egenskap} \\ 0 & \text{annars} \end{cases}$$

Statistikor 1

- Notera att dessa egenskaper är **funktioner** av de i stickprovet ingående observationerna
- Kallas **urvalskaraktistikor** eller med ett annat ord **statistikor**
- Medelvärdet är ett exempel på en **statistika**.
- Innan vi observerar stickprovet kan dessa **statistikor** betraktas som vadå?
- **Stokastiska variabler!**

Statistikor 2

- Stickprovsmedelvärdet som s.v. kan betecknas med versal $\bar{X}$
- Stickprovsvariansen likaså S^2

$\bar{X}$ och S^2 har sina respektive

- **väntevärden** $E(\bar{X})$ och $E(S^2)$
- **varianser** $V(\bar{X})$ och $V(S^2)$

och

- **fördelningar** $f_{\bar{X}}(\bar{x})$ och $f_{S^2}(s^2)$

Enstaka observationer

Antag att samtliga s.v. i urvalet

$$X_1, X_2, \dots, X_n$$

har samma väntevärde och varians

$$E(X_i) = \mu \text{ och } V(X_i) = \sigma^2$$

för alla $i = 1, 2, \dots, n$.

Observera att vi använder symbolerna μ och σ^2 för att slippa skriva $E(X_i)$ och $V(X_i)$ varje gång.

Stickprovsmedelvärdet 1

- **Väntevärde:**

$$\begin{aligned} E(\bar{X}) &= E\left[\frac{X_1 + X_2 + \dots + X_n}{n}\right] \\ &= \frac{1}{n} [E(X_1) + E(X_2) + \dots + E(X_n)] \\ &= \frac{1}{n} \cdot n \cdot E(X_i) = \mu \end{aligned}$$

Stickprovsmedelvärdet 2

- Vi antar att samtliga X_i är korsvis oberoende sinsemellan

- **Varians:**

$$\begin{aligned} V(\bar{X}) &= V\left[\frac{X_1 + \dots + X_n}{n}\right] \\ &= V\left(\frac{X_1}{n}\right) + \dots + V\left(\frac{X_n}{n}\right) \\ &= \frac{1}{n^2} V(X_1) + \dots + \frac{1}{n^2} V(X_n) \\ &= \frac{1}{n^2} \cdot n \cdot V(X_i) = \frac{\sigma^2}{n} \end{aligned}$$

Simultanfördelningen

- Vi kommer ihåg: om X_1 och X_2 är två s.v. med resp. marginalfördelningar $f_{X_1}(x_1)$ och $f_{X_2}(x_2)$ och om

$$f(x_1, x_2) = f_{X_1}(x_1) \cdot f_{X_2}(x_2)$$

så är X_1 och X_2 oberoende

- Med ett helt urval om n stycken s.v. gäller motsvarande, om

$$f(x_1, \dots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n)$$

så är $X_1, \dots, X_n$ korsvis oberoende

Exempel

- Antag att $X_i, i = 1, \dots, 3$ är oberoende diskreta s.v. med gemensam samma frekvensfunkt.

$$f(x_i) = 1/3, x_i = 1, 2, 3$$

- Definiera Y som snittet av dessa, dvs. $Y = X_1 + X_2 + X_3$.

- **Vad har $Y/3 = X$ -bar för fördelning?**

- Vi tittar på antalet möjliga utfall. Enligt multiplikationsprincipen får vi $3^3 = 27$ möjliga utfall/urval

Exempel, forts.

27 möjliga stickprov

x_1	1	1	1	2	1	1	3	1	2
x_2	1	1	2	1	1	3	1	2	1
x_3	1	2	1	1	3	1	1	2	2
Snitt	1	4/3	4/3	4/3	5/3	5/3	5/3	5/3	5/3
X_1	2	1	1	2	2	3	3	2	3
X_2	2	2	3	1	3	1	2	2	3
X_3	1	3	2	3	1	2	1	2	1
Snitt	5/3	2	2	2	2	2	2	2	7/3
X_1	3	1	2	2	3	3	3	2	3
x_2	1	3	2	3	2	3	2	3	3
x_3	3	3	3	2	2	2	3	3	3
Snitt	7/3	7/3	7/3	7/3	7/3	8/3	8/3	8/3	3

Frekvensfunktionen för snittet (x -bar)

$\bar{x}$	1	4/3	5/3	2	7/3	8/3	3
$f(\bar{x})$	1/27	3/27	6/27	7/27	6/27	3/27	1/27

Exempel, forts.

- Beräkna väntevärde och varians för X_i

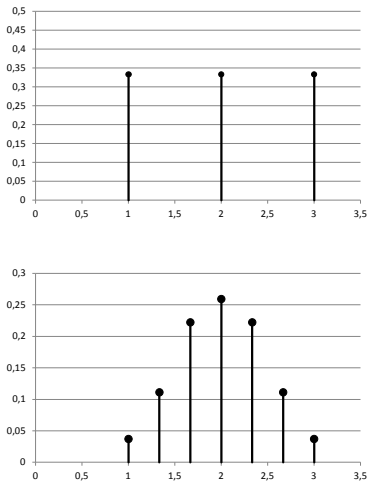
$$E(X_i) = 2 \quad V(X_i) = 2/3 \approx 0,667$$

- Beräkna väntevärde och varians för $\bar{X}$

$$E(\bar{X}) = 2 \quad V(\bar{X}) = (2/3)/3 = 2/9 \approx 0,222$$

- **Rita frekvensfunktionen både för X_i och för X -bar!**

Exempel, forts.



Stickprovsmedelvärdet 3

- Nu vet vi väntevärde och varians för stickprovsmedelvärdet!
- I exemplet innan var utfallsrummet för X_i diskret, $\Omega_X = \{1, 2, 3\}$
- För $\bar{X}$ var utfallsrummet också diskret, $\Omega_{\bar{X}} = \{1, 4/3, 5/3, 2, 7/3, 8/3, 3\}$

Frågor som uppstår

- Vad gäller i andra fall?
- Om X_i 'na är normalfördelade?
- Eller när $n \rightarrow \infty$?

Fall 1 (avsnitt 15.2)

- Observationer från en normalfördelad population
- Variansen σ^2 är känd
- **Resultat:** Om alla X_i dessutom är normalfördelade med samma väntevärde och varians så är $\bar{X}$ också normalfördelad
- Dvs. $X_i \sim N(\mu, \sigma^2) \Rightarrow \bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

Fall 1, forts.

Vi kommer ihåg transformationen

$$Z = \frac{X - E(X)}{\sqrt{V(X)}} = \frac{X - \mu}{\sigma}$$

och att $Z \sim N(0, 1)$

För $\bar{X}$ har vi motsvarande transformation

$$Z = \frac{X - E(\bar{X})}{\sqrt{V(\bar{X})}} = \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

och att $Z \sim N(0, 1)$

Fall 1: Exempel

Antag $X_i \sim N(40,16)$ för $i = 1, \dots, 16$.

- Beräkna $P(X_i > 42)$

$$P(X > 42) = P\left(\frac{X-40}{\sqrt{16}} > \frac{42-40}{\sqrt{16}}\right) = P(Z > 0,5) \\ = 1 - P(Z \leq 0,5) = 1 - 0,6915 = 0,3085$$

- Beräkna $P(\bar{X} > 42); n = 16$

$$P(\bar{X} > 42) = P\left(\frac{\bar{X}-40}{\sqrt{16/16}} > \frac{42-40}{\sqrt{16/16}}\right) = P(Z > 2) \\ = 1 - P(Z \leq 2) = 1 - 0,97725 = 0,02275$$

Fall 1, forts.

- Vi inser att fördelningen för $\bar{X}$ är **smalare** än den ursprungliga fördelningen
- Dvs. den har en **mindre varians**
- Variansen beror på n ; ju större n desto mindre varians.
- När $n \rightarrow \infty$ följer att $V(\bar{X}) \rightarrow 0$
- Kom även ihåg att $E(\bar{X}) = \mu$

χ^2 -fördelningen 1

För normalfördelade X_i har vi

$$Z_i = \frac{X_i - \mu}{\sigma}$$

där $Z_i \sim N(0,1)$

Bilda kvadraten

$$Z_i^2 = \frac{(X_i - \mu)^2}{\sigma^2}$$

och summera n st oberoende Z_i^2 över stickprovet

$$\chi^2 = \sum_{i=1}^n \frac{(X_i - \mu)^2}{\sigma^2}$$

χ^2 -fördelningen 2

- χ^2 är en **stokastisk variabel**
- χ^2 är χ^2 -fördelad med n **frihetsgrader** ← parameter!
- Notera risken för förvirring;
 χ^2 används som symbol både för den stokastiska variabeln och dess fördelning!
- Vi skriver $\chi^2 \sim \chi^2(n)$
- Om ni vill undvika förvirring skriv t.ex. Q^2 isf χ^2 och $Q^2 \sim \chi^2(n)$

χ^2 -fördelningen 3

Anta att $\chi^2 \sim \chi^2(n)$. Då gäller att

- Utfallsrummet för χ^2 är $(0, \infty)$
- $E(\chi^2) = n$
- $V(\chi^2) = 2n$ ↑
Rimligt?
- Obs! χ^2 -fördelningen är inte symmetrisk
- När vi använder tabellen måste vi ibland leta upp ett värde för vänstersidan och ett annat för högersidan

χ^2 -fördelningen 4

Stickprovsvarians:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Transformation (sid 14):

$$C^2 = \frac{(n-1)S^2}{\sigma^2}$$

Det gäller att $C^2 \sim \chi^2(n-1)$ och

- $E(C^2) = n - 1$
- $V(C^2) = 2(n - 1)$

t -fördelningen 1

Vi skapar ytterligare en stokastisk variabel ur några som vi redan har!

- $Z \sim N(0,1)$
- $\chi^2 \sim \chi^2(v)$
- Z och χ^2 är oberoende (viktigt)

Skapa den nya s.v. T enligt

$$T = \frac{Z}{\sqrt{\chi^2/v}}$$

t -fördelningen 2

- T är t -fördelad med v ↑ Parameter! **frihetsgrader**
- Vi skriver $T \sim t(v)$ ↑ Rimligt?
- Utfallsrummet för T är $(-\infty, \infty)$
- Om $v > 1$, $E(T) = 0$
- Om $v > 2$, $V(T) = v/(v-2)$ ← $v \rightarrow \infty$?
- t -fördelningen påminner om standardnormalfördelningen (Z)
– tryck bara till på toppen och lite sannolikhet rinner ut åt sidorna!

t-fördelningen 3

- t-fördelningen är liksom standard-normalfördelningen symmetrisk kring noll (0)
- När vi använder tabellen räcker det att slå upp värdet för högersidan av fördelningen och utnyttja

$$P(T \leq -t_\alpha) = P(T > t_\alpha)$$

- Dubbelbokning på schemat den 19 oktober; både genomgång av tenta samt gruppövning för gr. B
- **Nytt datum för den blir 17 oktober kl. 16 i B413.**
- Uppdaterat schema finns på hemsidan

Repetition

F4 Samplingfördelningar

- **Urval** eller **stickprov** som en uppsättning s.v.:

$$X_1, X_2, \dots, X_n$$

- **Observationer** av dessa s.v.:

$$x_1, x_2, \dots, x_n$$

- **Egenskaper** hos stickprovet kan bl.a. vara:
 - stickprovsmedelvärdet, $\bar{X}$
 - stickprovsvariansen, S^2
 - andelen i stickprovet, p eller P

Statistikor!

Repetition

Statistikor

$\bar{X}$ och S^2 (och p) har sina respektive

- **väntevärden** $E(\bar{X})$ och $E(S^2)$
 - **varianser** $V(\bar{X})$ och $V(S^2)$
- och
- **fördelningar** $f_{\bar{X}}(\bar{x})$ och $f_{S^2}(s^2)$

Repetition

X-bar

Antag att alla s.v. i urvalet $X_1, \dots, X_n$ har samma väntevärde och varians;
 $E(X_i) = \mu$ och $V(X_i) = \sigma^2$.

Antag att samtliga $X_1, \dots, X_n$ är kors-vis oberoende.

Då gäller:

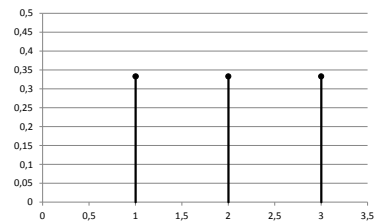
$$E(\bar{X}) = E(X_i) = \mu$$

$$V(\bar{X}) = \frac{V(X_i)}{n} = \frac{\sigma^2}{n}$$

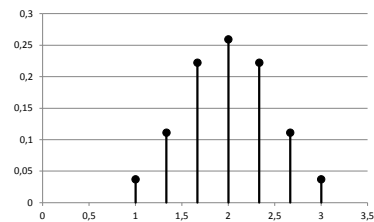
Repetition

Samplingfördelningar

Exemplet: de enskilda observationernas fördelning



Exemplet: stickprovssnittets fördelning

*Repetition*

Fall 1 (avsnitt 15.2)

- Observationer från en normalfördelad population (oberoende)
- Variansen σ^2 är känd
- **Resultat:** Om alla X_i dessutom är normalfördelade med samma väntevärde och varians så är $\bar{X}$ också normalfördelad
- Dvs. $X_i \sim N(\mu, \sigma^2) \Rightarrow \bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

Repetition

χ^2 -fördelningen

Antag n stycken oberoende Z_i sådana att alla $Z_i \sim N(0,1)$.

Summan av alla dessa, $\chi^2 = \sum_{i=1}^n Z_i^2$

är χ^2 -fördelad med n **frihetsgrader** och vi skriver $\chi^2 \sim \chi^2(n)$.

- Utfallsrummet = $(0, \infty)$
- $E(\chi^2) = n$
- $V(\chi^2) = 2n$
- fördelningen är inte symmetrisk

Repetition

Fördelningen för S^2

Stickprovsvariansen:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Transformation (sid 14):

$$C^2 = \frac{(n-1)S^2}{\sigma^2}$$

Det gäller att $C^2 \sim \chi^2(n-1)$ och

- $E(C^2) = n - 1$
- $V(C^2) = 2(n - 1)$

Repetition

t-fördelningen

Vi skapar ytterligare en s.v.:

- $Z \sim N(0,1)$
- $\chi^2 \sim \chi^2(v)$, dvs. v frihetsgrader
- Z och χ^2 är oberoende (viktigt)

Skapa den nya s.v. T enligt $T = \frac{Z}{\sqrt{\chi^2/v}}$

T är t-fördelad med v **frihetsgrader** och vi skriver $T \sim t(v)$

- Utfallsrummet för T är $(-\infty, \infty)$
- Om $v > 1$, $E(T) = 0$
- Om $v > 2$, $V(T) = v/(v-2)$

Fall 2 (avsnitt 15.5)

- Observationer från en normal-fördelad population (oberoende)
- Variansen σ^2 är okänd

För X hade vi transformationen

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = Z \sim N(0,1)$$

Oftast är σ^2 okänd; skapa istället

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} = T \sim ?$$

Fall 2, forts.

Från sid 19, för att ta reda på fördelningen för T :

$$\begin{aligned} T &= \frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}}{\frac{S/\sqrt{n}}{\sigma/\sqrt{n}}} = \frac{Z}{S/\sigma} = \frac{Z}{\sqrt{S^2/\sigma^2}} \\ &= \frac{Z}{\sqrt{\frac{(n-1)S^2}{\sigma^2}/(n-1)}} = \frac{Z}{\sqrt{C^2/(n-1)}} \end{aligned}$$

där $Z \sim N(0,1)$ och $C^2 \sim \chi^2(n-1)$

$$\Rightarrow T \sim t(n-1)$$

Exemplet från Fall 1

Antag $X_i \sim N(\mu, \sigma^2)$ för $i = 1, \dots, 16$.

- Beräkna $P(\bar{X} > \mu + 2)$; $n = 16$, men μ, σ^2 är okända, använd $s^2 = 16$

$$P(\bar{X} > \mu + 2) = P(\bar{X} - \mu > 2)$$

$$= P\left(\frac{\bar{X} - \mu}{\sqrt{s^2/n}} > \frac{2}{\sqrt{s^2/n}}\right) = P\left(T > \frac{2}{\sqrt{16/16}}\right)$$

$$= P(T > 2) = 0,031973$$

där $T \sim t(15)$

*Jämför med fallet med
känd varians; gav
sannolikheten 0,02275*

*Obs! Tabellen är inte användbar här! Har
använt program för att beräkna sannolikheten.*

Sammanfattning

n stycken **oberoende** observationer från en **normalfördelad** population med väntevärde μ och varians σ^2 :

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \quad Z \sim N(0,1)$$

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \quad T \sim t(n-1)$$

$$C^2 = \frac{(n-1)S^2}{\sigma^2} \quad C^2 \sim \chi^2(n-1)$$

Vad som gäller för $\bar{X}$

- **Alltid:** $E(\bar{X}) = E(X_i) = \mu$

- **Oberoende** observationer:

$$V(\bar{X}) = \frac{V(X_i)}{n} = \frac{\sigma^2}{n}$$

- **Normalfördelade** observationer:

$$Z \sim N(0,1) \text{ alt. } T \sim t(n-1)$$

- Ej normalfördelade observationer men $n \rightarrow \infty$

$$Z \rightarrow N(0,1)$$

Centrala gränsvärdessatsen

- Från sid 20:

Formen på samplingfördelningen för en Z-transformation av urvalsmedelvärdet närmar sig formen för en standardiserad normalfördelning då urvalsstorleken ökar.

- Dvs.

$$Z = \frac{\bar{X} - E(X)}{V(X)/\sqrt{n}} \rightarrow N(0,1)$$

när $n \rightarrow \infty$

CGS

- Gäller oavsett vilken fördelning de enskilda observationerna är dragna från.

- Tumregel:**

minst $n = 30$ observationer

Men ...

- Ju mer populationens fördelning skiljer sig från en normalfördeln., desto större stickprov krävs.
– min erfarenhet, CGS fungerar väldigt bra med **symmetriska** fördelningar

CGS

- Vi kan använda CGS direkt på $\bar{X}$ eller linjära funktioner av $\bar{X}$

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \rightarrow N(0,1)$$

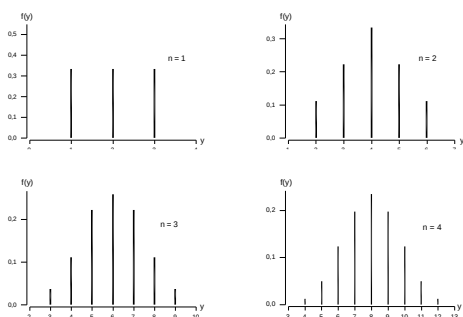
$$\bar{X} = \frac{\sigma}{\sqrt{n}} Z + \mu \rightarrow N(\mu, \frac{\sigma^2}{n})$$

$$T_x = n\bar{X} \rightarrow N(n\mu, n\sigma^2)$$

(Kap 15 sid 23-24)

CGS

- När $n \rightarrow \infty$



Andelar/proportioner 1

- Andelen i stickprovet med ngn egenskap:

$$p = \frac{1}{n} \sum_{i=1}^n x_i \quad x_i = \begin{cases} 1 & \text{om i:haregenskape} \\ 0 & \text{annars} \end{cases}$$

- Detta är inget annat än ett medelvärde! Alltså CGS!

- Beteckna med P den stokastiska variabeln

$$P = \frac{1}{n} \sum_{i=1}^n X_i$$

Andelar/proportioner 2

- Antag att vi gör n oberoende observationer
- Antag att den sanna andelen/proportionen är π där $0 < \pi < 1$
- Låt $Y = n \cdot P$ vara antalet i urvalet med egenskapen
- **Vad har Y för fördelning?**

$$Y \sim \text{Bin}(n, \pi)$$

med väntevärde och varians

$$E(Y) = n\pi \quad V(Y) = n\pi(1-\pi)$$

Andelar/proportioner 3

- Transformera Y tillbaks till P

$$E(P) = E(Y/n) = \pi$$

$$V(P) = V(Y/n) = \pi(1-\pi)/n$$

- **Men P är ett medelvärde och enligt CGS**

$$P \rightarrow N(\pi, \frac{\pi(1-\pi)}{n}) \quad \text{när } n \rightarrow \infty$$

Exempel

- Antag att 52 % skulle rösta på republikanerna i det kommande presidentvalet i USA
- Vi tar ett urval av storlek $n = 100$
- **Vad är sannolikheten att $P \leq 0,50$?**

$$\text{• Definiera } Z = \frac{P - 0,52}{\sqrt{\frac{0,520,48}{100}}} = \frac{P - 0,52}{0,04996}$$

och enligt CGS så $Z \rightarrow N(0,1)$

Exempel, forts.

$$\text{Så } P(Z \leq \frac{0,50 - 0,52}{\sqrt{\frac{0,520,48}{100}}}) = P(Z \leq -0,40)$$

$$= 1 - \Phi(0,40) = 0,34446$$

Men bättre med **halvkorrektion**

$$P(Z \leq \frac{0,50 - 0,52 + \frac{1}{200}}{\sqrt{\frac{0,520,48}{100}}}) = P(Z \leq -0,30)$$

$$= 1 - \Phi(0,30) = 0,38200$$

Exakt beräkning ger 0,38162

Avslutning CGS

X är typiskt baserad på en **summa** av oberoende lika fördelade s.v.

- Kom ihåg:

$$Z = \frac{\bar{X} - E(X)}{V(X)/\sqrt{n}} \rightarrow N(0,1)$$

när $n \rightarrow \infty$

Egentligen säger CGS att **summor** av olf s.v. går mot en normalfördelning.

Väldigt många **statistikor** bygger på summor av olf s.v. Hur är det med stickprovsvariansen t.ex.?

F5 CGS rep

Repetition

- Från sid 20:

Formen på samplingfördelningen för en Z-transformation av urvalsmedelvärdet närmar sig formen för en standardiserad normalfördelning då urvalsstorleken ökar.

- Dvs. när $n \rightarrow \infty$ $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \rightarrow N(0,1)$

- Gäller oavsett vilken fördelning de enskilda observationerna är dragna från.

- Tumregel:** minst $n = 30$ observationer

CGS

Repetition

- Vi kan använda CGS direkt på $\bar{X}$ eller linjära funktioner av $\bar{X}$

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \rightarrow N(0,1)$$

$$\bar{X} = \frac{\sigma}{\sqrt{n}} Z + \mu \rightarrow N(\mu, \frac{\sigma^2}{n})$$

$$T_x = n\bar{X} \rightarrow N(n\mu, n\sigma^2)$$

(Kap 15 sid 23-24)

Utveckla och använd CGS

Repetition

Exempel 1:

- Antag att $X_i \sim Po(\lambda)$
- W = summan av n oberoende X_i
- Då är $W \sim Po(n\lambda)$
- Men enligt CGS: $W \rightarrow N(n\lambda, n\lambda)$

Exempel 2:

- Antag att $Z_i \sim N(0,1)$
- Då gäller att $Z_i^2 \sim \chi^2(1)$
- W = summan av n oberoende Z_i^2
- Då är $W \sim \chi^2(n)$
- Men enligt CGS: $W \rightarrow N(n, 2n)$

Kap 16 Uppskattning

- Ofta vill man kunna säga något om ett **populationsmedelvärde** μ (eller gruppmedelvärden)

alt.

- Utfallet av ett experiment tänks vara fördelat enligt någon modell och vi vill kunna säga ngt om **väntevärdet** $E(X_i) = \mu$
- Den enda information vi har är **stickprovet** och de **statistikor** vi kan beräkna

Uppskattning

Vissa statistikor känns mer naturliga att använda som uppskattningar för olika parametrar än andra.

Ex. stickprovs-

- -medelvärdet $\bar{X}$ skattar μ
- -variansen S^2 skattar σ^2
- -andelen P skattar π

Kallas för **punktskattningar**.

- det enskilda värde som man använder som bästa gissning för värdet på den okända parametern

Begrepp och termer

- **Uppskattning** eller bara **skattning**
- Statistikan som används för att skatta parametern, i detta fall μ , kallas en **skattning** av μ
– (eng. *estimator*)
- Det numeriska värde som man sedan observerar kallas en **skattning** av μ
– (eng. *estimate*)

Osäkerhet i skattningar 1

Definitionen av begreppet **osäkerhet** på sid 5 ger en god sammanfattning:

- När sannolikheten är stor att skattningen ligger långt från det sanna värdet så är osäkerheten stor.
- När sannolikheten är liten att skattningen ligger långt från det sanna värdet så är osäkerheten liten.

Osäkerhet i skattningar 2

En **statistika** som väljs för att skatta en parameter är en s.v.

- Om denna har en **stor varians** så är det större sannolikhet att den ska hamna "långt bort"
 - dvs. **stor osäkerhet**
- Om den har en **liten varians** så är sannolikheten relativt sett mindre att den ska hamna "långt bort"
 - dvs. **liten osäkerhet**

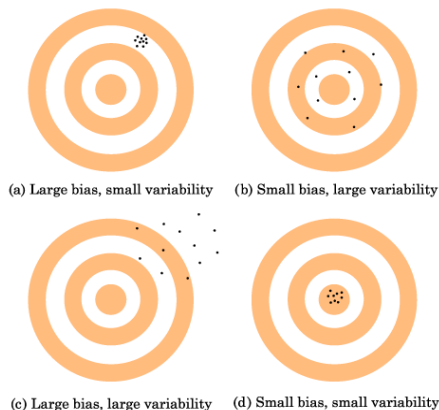
• **Jämför med Figur 16.1 sid 5**

Skattningens egenskaper

- **Väntevärdesriktig:** $E(\bar{X}) = \mu$
 - I det långa loppet kommer vi i snitt pricka rätt.
 - **Bias** och **unbiased**
- **Konsistens:** $V(\bar{X}) = \sigma^2 / n$
 - När $n \rightarrow \infty$ går variansen mot noll
 - Betyder att sannolikheten att ligga långt från μ minskar när n ökar
 - Osäkerheten minskar.

Skatta μ

- Väntevärdesriktig (eng. **unbiased**) och osäkerhet



Skatta μ

- Antag att vi drar ett stickprov av storlek n från en normalfördelad population med känd varians σ^2
- Vi använder $\bar{X}$ som punktskattning för μ
- Vi vill **uttala oss om osäkerheten** kring punktskattningen.

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0, 1)$$

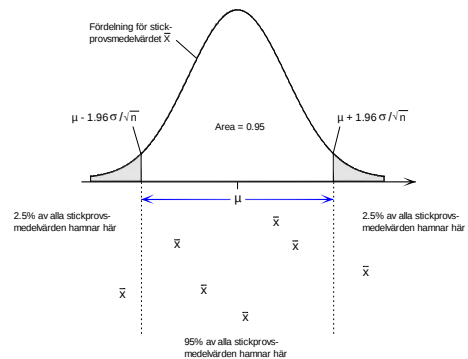
Sannolikheten att ligga i ett intervall

- Sannolikheten att $\bar{X}$ hamnar i intervallet $(\mu - 1,96 \cdot \frac{\sigma}{\sqrt{n}}, \mu + 1,96 \cdot \frac{\sigma}{\sqrt{n}})$

$$\begin{aligned}
 &P(\mu - 1,96 \cdot \frac{\sigma}{\sqrt{n}} < \bar{X} < \mu + 1,96 \cdot \frac{\sigma}{\sqrt{n}}) \\
 &= P(-1,96 < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < 1,96) \\
 &= P(-1,96 < Z < 1,96) \\
 &= P(Z < 1,96) - P(Z < -1,96) \\
 &= 2 \cdot \Phi(1,96) - 1 \quad (\text{se även Tabell 2}) \\
 &= 2 \cdot 0,975 - 1 = 0,95
 \end{aligned}$$

Sannolikheten att ligga i ett intervall

- Upprepade stickprov ger olika medelvärden
- 95 % sannolikhet att $\bar{X}$ hamnar i intervallet



Lägg intervallet runt det observerade $\bar{X}$ istället!

Vänd på det

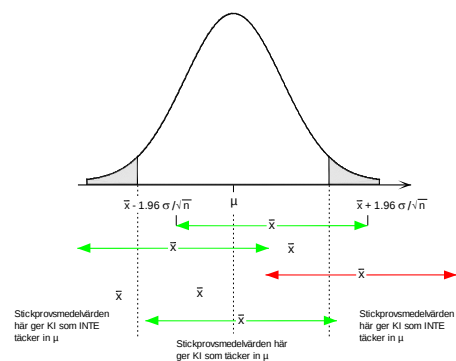
- Byt plats på $\bar{X}$ och μ :
- Dvs. låt **gränserna** för intervallet vara **slumpmässiga**

$$\begin{aligned}
 &P(\mu - 1,96 \cdot \frac{\sigma}{\sqrt{n}} < \bar{X} < \mu + 1,96 \cdot \frac{\sigma}{\sqrt{n}}) \\
 &= P(-\bar{X} - 1,96 \cdot \frac{\sigma}{\sqrt{n}} < -\mu < -\bar{X} + 1,96 \cdot \frac{\sigma}{\sqrt{n}}) \\
 &= P(\bar{X} + 1,96 \cdot \frac{\sigma}{\sqrt{n}} > \mu > \bar{X} - 1,96 \cdot \frac{\sigma}{\sqrt{n}}) \\
 &= P(\bar{X} - 1,96 \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1,96 \cdot \frac{\sigma}{\sqrt{n}}) \\
 &= 0,95
 \end{aligned}$$

visste vi redan

Stokastiska intervall

- 95 % av alla möjliga intervall täcker μ
- Man skattar ett intervall, inte bara en punkt



Jämför med Figur 16.2 sid 8

Konfidensintervall 1

- När vi observerat ett $\bar{x}$ och skapar ett KI så är det inte stokastiskt
- Det är m.a.o. inte korrekt att säga att det är 95 % sannolikhet att intervallet täcker μ
- Men om vi upprepar hela proceduren skulle 95 % av alla KI täcka in μ
- **Konfidensgraden** är 95 %
 - "graden av tilltro"
- Kallas ett **95 % konfidensintervall för μ**
- Längden på intervallet kallas den **statistiska felmarginalen**

Konfidensgraden

Varför just 95 %?

- Varför inte 50, 75, 90 eller 99,9 %?
- Man kan välja konfidensgraden helt fritt men 95 % är vanligt om inte det vanligaste.

Hur säker vill man vara?

- Bredare intervall → högre konfidens
 - fler KI täcker μ
- Högre konfidens → bredare intervall
 - Vad krävs för 100 % konfidens?
 - Ett oändligt intervall $(-\infty, \infty)$

Konfidensintervall 2

Notera att för givet värde på σ^2 så bestäms intervallets längd, dvs. felmarginalen, av

- **Konfidensgraden**
 - 95 % ger oss värdet 1,96 (Tabell 2)
 - Vilket värde skulle vi få vid 90 %?
 - 99 %?
- **Stickprovsstorleken**
 - eftersom $V(\bar{X}) = \frac{\sigma^2}{n}$
 - större stickprov → mer information

Konfidensgraden

- Bestäm konfidensgrad = $1-\alpha$
 - Ex. 95 % $\Rightarrow 0,95 = 1-\alpha \Rightarrow \alpha = 0,05$
- Hälften av α läggs i vardera svans
 - Se Figur 16.3 sid 10
- Slå upp värdet för $z_{\alpha/2}$ i Tabell 2
 - Ex. $z_{0,025} = 1,96 \Rightarrow 95 \% \text{ KI}$
 - Ex. $z_{0,05} = 1,6449 \Rightarrow 90 \% \text{ KI}$
 - Ex. $z_{0,005} = 2,5758 \Rightarrow 99 \% \text{ KI}$

Formler K.I. för μ

1. Normalfördelade observationer, variansen σ^2 är känd

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \quad \text{Tabell 2}$$

2. Normalfördelade observationer, variansen σ^2 är okänd; använd s^2

$$\bar{x} \pm t_{\alpha/2}^{(n-1)} \cdot \frac{s}{\sqrt{n}} \quad \text{Tabell 3}$$

Formler K.I. för μ

3. Ej normalfördelning men stort stickprov, variansen σ^2 är känd

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \quad \text{Tabell 2}$$

4. Ej normalfördelning men stort stickprov, variansen σ^2 är okänd

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}} \quad \text{Se till att ha stort } n!$$

Exempel

Antag att vi observerar följande

$$\bar{x} = 33, s^2 = 16, n = 25$$

Bestäm ett 95 % KI för μ

- $0,95 = 1 - \alpha \Rightarrow \alpha/2 = 0,025$
- Frihetsgrader = $n - 1 = 24$
- I Tabell 3 avläses $t_{0,025}^{(24)} = 2,064$
- Ett 95 % KI för μ ges av

$$\bar{x} \pm t_{0,025}^{(n-1)} \cdot \frac{s}{\sqrt{n}} = 33 \pm 2,064 \cdot \frac{4}{\sqrt{25}}$$

$$= 33 \pm 1,6512 \text{ eller } (31,35, 34,65)$$

Bredvidläsning

- Nyquist: Kap 4 och 10
- Thurén: Kap 5
- Dessa avsnitt kommer inte att examineras explicit men det rekommenderas att ni läser igenom dem